Bayesian Variable Selection for Quantile Regression

Dilani Kaveendri , Prof. Sally Cripps and Dr. Nandini Ramesh

November 29, 2022

- Quantile Regression: Modelling various percentiles of the distributions and thus getting a more complete picture of the set.

- **Bayesian Quantile Regression**: framework to deal with both parameter estimation and uncertainty quantification.

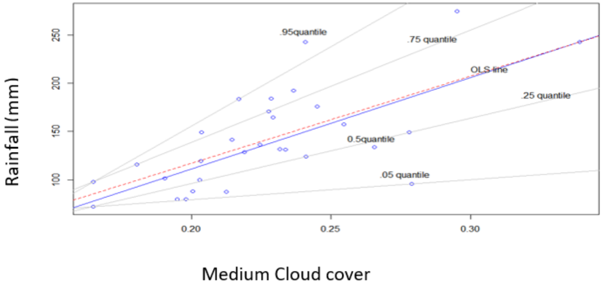


Figure: Quantile distribution of Rainfall vs Medium cloud cover for Sydney Observatory Hill

Why Bayesian Variable selection?

- Allows us to make inferences on the dependency between potential predictive factors on the tail, or extreme, of the response distribution while quantifying uncertainty.

- Priors : **Spike and Slab**, **Bayesian LASSO, Adaptive Lasso , Horseshoe**

- Application : **Rainfall** is an aggregate of different climate phenomena

- Variables : Climate Indices (Southern Oscillation Index (SOI),Dipole Mode Index (DMI), Southern Annular Mode (SAM) and all interactions)

Application

Rainfall data from 1950-2021 (72 years) for Sydney Observatory Hill.

March - One of the Rainiest month

- **SOI** : Measures El Niño and La Niña events
- **DMI**: Measures the Indian Ocean Dipole
- **SAM**: Measures the Southern Annular Mode

General Model for a given site (s) and for time (t):

$$y_t = X_t \beta_t^q + \epsilon_t$$

ϵ_t distributed as **Asymmetric Laplace (ALD)**

The q th regression quantile $(0 < q < 1)$ is defined as any solution, β^q, to the quantile regression minimisation problem

$$\min_{\beta} \sum_t \rho_q (y_t - X_t \beta_t^q)$$

where the loss function $\rho_q(u) = u(q - I(u < 0))$

ALD can be written as (Kozumi and Koabayshi (2009)),

$$y_t = X_t\beta_t^q + \sigma\theta z_t + \sigma\tau\sqrt{z_t}u_t$$

where $v_t = \sigma z_t$, $z_t \sim \exp(1)$ and $u_t \sim N(0, 1)$

Likelihood can be rewritten as ,

$$f(y|\beta^q, v, \sigma) \propto (\prod_{t=1}^{T} (\sigma v_t)^{\frac{-1}{2}}) \exp\{-\sum_{t=1}^{T} \frac{(y_t - X_t\beta_t^q - \theta v_t)^2}{2\tau^2\sigma v_t}\}(1)$$

Prior: 1). Spike and Slab prior in QR (Sampling model and parameter space)

- $X_t = (\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p)$ and $\mathbf{x}_k = (x_{1k}, \ldots, x_{nk})$
- Consists of global climate indicators, SOI, SAM, DMI and their interactions at time t
- γ_{kt}^q, where, $\gamma_{kt}^q = 1$ if variable $\mathbf{x}_k$ is useful in prediction y_t^q, at time t, and $\gamma_{kt}^q = 0$ otherwise, for $k = 0, 1, \ldots, p$
- The posterior distribution of γ_t

$$\pi(\gamma_t|y) = \frac{f(y_t|\gamma_t)\pi(\gamma_t)}{f(y)}$$

  where $\mathbf{y} = (y_1, \ldots, y_n)$
- Select variables that appear at least in 50% of visited models (i.e Marginal Inclusion Probability > 0.5)

Prior: 1) Spike and Slab prior in QR (Sampling model and parameter space)

Define the set A_1 s.t $A_1 = \{k; \gamma_{kt} = 1\}$ and $||A_1|| = n_1$ and $||A_0|| = n_0$
For each model M_γ,

$$\beta^q_{A_1 t} | \gamma_t, \sigma \sim N(0, \sigma \lambda_k^{-1})$$

k = 1,.....,p and

$$\lambda_k \sim \text{Gamma}(1/2, 1/2)$$

$$\beta^q_{A_0 t} | \gamma_t, \sigma \sim \delta(0)$$

We assume γ are i.i.d. Bernoulli (Be) (π), so that

$$\gamma_t | \pi \sim Be(\pi)$$
$$\pi \sim \text{Beta}(a_1, a_2)$$
$$\sigma \sim \text{IG}(a, b), v_t \sim \exp(\sigma)$$

Prior : 2) Bayesian LASSO for QR (Estimating regression parameters)

Laplace prior

$$\pi(\beta_k|\lambda, \sigma) = \frac{\lambda}{2\sqrt{\sigma}} exp - \frac{\lambda|\beta_k|}{\sqrt{\sigma}}$$

with $\lambda \geq 0$

The above Laplace prior can be written in a hierarchical form,

$$\beta^q \sim N(0, \sigma D_s)$$

,
where $D_s = diag(s_1^2, ..., s_p^2)$ and p= no of coefficient parameters.

$$s_k^2 \sim \text{Gamma}(1, \lambda^2)$$
$$\lambda^2 \sim \text{Gamma}(a_1, b_1)$$

Prior: 3) Bayesian Adaptive LASSO for QR

This is a scale mixture of normal representation of the Laplace density

$$\pi(\beta_k|\lambda_k, \sigma) = \frac{\lambda_k}{2\sqrt{\sigma}} exp - \frac{\lambda_k |\beta_k|}{\sqrt{\sigma}}$$

Prior : 4) Horseshoe prior for QR

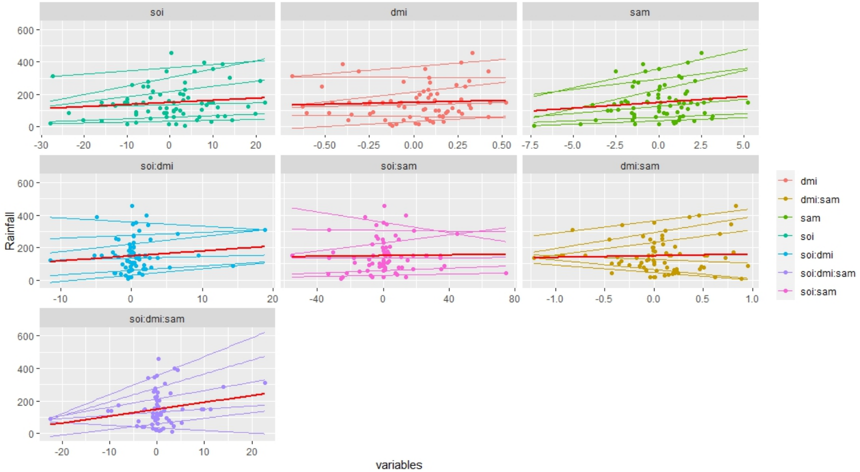
The general form of independent global-local prior takes the following hierarchical form:

$$\beta_k | \lambda_k^2, \nu^2 \sim N(0, \sigma \lambda_k^2 \nu^2)$$
$$\lambda_k^2 \sim C_+(0, 1)$$
$$\nu^2 \sim C_+(0, 1)$$

Results

- Using likelihood of the model and priors for different techniques conditional posterior distributions are derived.
- Based on posterior credible Intervals and MIPs for significance variables selected
- All methods were implemented using MCMC(Gibbs sampler) using 10,000 iterations with 2000 burn-ins in R (Without using inbuilt packages)

Results



Frequentist Quantile regression plot for Rainfall vs SOI,DMI and SAM with interactions for Sydney in March at q=0.1,0.2,0.5,0.85,0.9,0.95 and red line represent least square regression line

Results for spike and slab prior

Table: MLEs values of regression coefficients for Sydney in March at q = 0.1, 0.5 and 0.9, Red color represent the significance variables based on 95% CI

Coefficient	MLEs		
	0.1	0.5	0.9
soi	0.06	0.05	0.11
dmi	0.05	0.01	-0.02
sam	0.01	0.06	0.03
soi:dmi	-0.04	0.05	0.15
soi:sam	0.05	0.05	0.07
dmi:sam	-0.11	-0.01	0.16
soi:dmi:sam	-0.03	0.15	0.26

Table: Posterior mean and (Marginal Inclusion Probability (MIP)) of regression coefficients for Sydney in March using Spike and slab prior QR at q = 0.1, 0.5 and 0.9

Coefficient	Posterior mean and MIPs		
	0.1	0.5	0.9
soi	0.016 (0.369)	0.010 (0.127)	0.092 (0.655)
dmi	0.016 (0.308)	0.000 (0.053)	-0.010 (0.499)
sam	0.031 (0.375)	0.009 (0.128)	0.019 (0.472)
soi:dmi	-0.035 (0.436)	0.007 (0.103)	0.124 (0.747)
soi:sam	0.013 (0.311)	0.003 (0.057)	0.044 (0.480)
dmi:sam	-0.091 (0.592)	0.002 (0.069)	0.147 (0.881)
soi:dmi:sam	-0.015 (0.354)	0.137 (0.570)	0.234 (0.835)

Results for Sydney in March

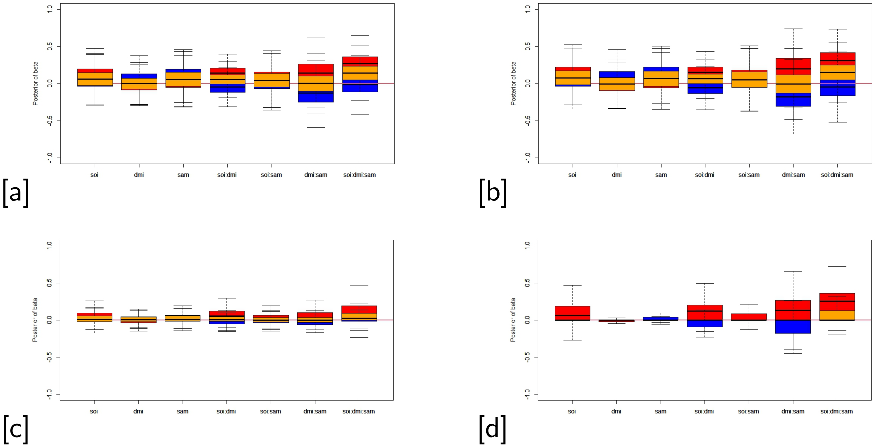


Figure: Box plots estimates for posterior iterates regression coefficients of soi,dmi,sam, soi:dmi, soi:sam, dmi:sam, soi:dmi:sam ([a] LASSO, [b] Adaptive LASSO, [c] Horseshoe, [d] Spike and slab), *blue* $->q=0.1$, *orange* $->0.5$ and *red* $->0.9$

Conclusion

- Spike and slab prior which is the "gold standard" for variable selection emulate with the frequentist results
- Spike and slab prior can be used as generalized method to make inferences on the significant variables and make predictions while quantify uncertainty
- Application wise it shows the importance use these methodology to model extreme rainfall

Future Research

Incorporating spatial dependence for spike and slab prior

- To induce the spatial dependency between the impact of a global climate indicator and rainfall quantile we place the following prior on γ^q_{kst}

$$\begin{aligned} \Pr(\gamma^q_{kst} = 1 | z_s) &= \pi^q_{kt}(z_s) \\ \pi^q_{kt}(z_s) &= \frac{\exp(g^q_t(z_s))}{1 + \exp(g^q_t(z_s))} \end{aligned}$$

where $z_s = (lat_s, lon_s)$ encodes the latitude and longitude at location s.

We place a Gaussian Process prior over the function $g^q_t(.)$

$$g^q_t \sim GP(\mu^q_t, \Omega^q_t)$$

and use the reproducing kernel Hilbert space defined by a two-dimensional thin plate Gaussian process prior to construct Ω^q_t.